PRICING AND THE ALLOCATION OF DATA PROCESSING RESOURCES

William Frank Habermann

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

PRICING AND THE ALLOCATION OF
DATA PROCESSING RESOURCES

by

William Frank Habermann

June 1975

Thesis Advisor: G. G. Brown

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Pricing and the Allocation of Data Processing Resources		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; June 1975
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William Frank Habermann		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE June 1975
		13. NUMBER OF PAGES 73
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

** Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Average Cost; Flat Rate; Flexible Pricing; Hard Money;

Soft Money; Charge Back

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis outlines the growing financial commitment by organizations to provide data processing services. The supply and demand characteristics of computers and microeconomics, pricing and accounting theories are reviewed in relation to resource allocation. The impact of no charge-back and charge-back policies on computer utilization are discussed in depth. Pricing goals--cost recovery and/or resource allocation, pricing principles and pricing structure--length of costing period, selecting

resources to be priced and quality of service are also investigated in depth. Suggested resource utilization measures are included in the Appendices. The advantages and disadvantages of average cost, flat rate and flexible pricing charge-back techniques are critically examined. Several examples of charge-back techniques utilized by commercial, educational and military organizations are presented. A time phased implementation plan of a charge-back system, beginning with average cost followed by flexible pricing, is recommended. A bibliography is included.

Pricing and the Allocation of
Data Processing Resources

by

William Frank Habermann
Lieutenant Commander, Supply Corps, United States Navy
B.S.B.A., University of Florida, 1962

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1975

ABSTRACT

This thesis outlines the growing financial commitment by organizations to provide data processing services. The supply and demand characteristics of computers and micro-economics, pricing and accounting theories are reviewed in relation to resource allocation. The impact of no charge-back and charge-back policies on computer utilization are discussed in depth. Pricing goals--cost recovery and/or resource allocation, pricing principles and pricing structure--length of costing period, selecting resources to be priced and quality of service are also investigated in depth. Suggested resource utilization measures are included in the Appendices. The advantages and disadvantages of average cost, flat rate and flexible pricing charge-back techniques are critically examined. Several examples of charge-back techniques utilized by commercial, educational and military organizations are presented. A time phased implementation plan of a charge-back system, beginning with average cost followed by flexible pricing, is recommended. A bibliography is included.

TABLE OF CONTENTS

I. INTRODUCTION

Over ten years ago, when some 16,000 computers were in use, Fortune magazine editorialized: "Today, men who ought to know say that the computer will have an effect on human life comparable to the invention of the steam engine. If that is only half true, anybody who hopes to play a responsible part in the second half of the twentieth century had better understand as much as he can of what computers do for-and-to-people." [Fortune, 1964]

This technological advance is important to today's manager not only because of the remarkable extension of man's ability but also because of the rapid growth and magnitude of the investment in the data processing resource. In the context of this paper, the data processing resource includes computer hardware, software, supplies and all the personnel who are employed as a result of the presence of the computer.

In the short space of thirteen years (1960-1973) the number of computers in the federal government has grown from 531 to 7,149. Excluded are analog and digital computers which are built or modified to special government design and integral to a weapons or space system. Of the 1973 total, the Departments of Defense and the Navy had 53% and 15.9%, respectively, of the government's installed computers. The total purchase price of system hardware in use by the federal government as of 30 June 1973 approximated $3.46 billion.

It should also be noted that the share of owned systems increased steadily and was over 81% by June 1973. [GSA, 1973]

Datamation surveys of approximately 200 commercial and governmental data processing installations in the United States and Canada produced the following breakdown of planned data processing expenditures for 1974 and 1975:

	Total	Personnel	Hardware	Supplies	Other
1974	$28.1 billion	46.9%	39.1%	6.8%	7.2%
1975	$31.8 billion	51.1%	37.8%	7.3%	3.8%

Planned annual total budget increases were 11%, 8% and 13%, respectively, for 1973-75.

These dollar amounts are impressive but inflation and shortages may account for much of the increases. Supplies and accessories cannot be overlooked as they were budgeted for a 1974 increase of 13% and a 1975 increase of 25%, mostly due to increases in forms and card stock. Additional attention must also be given to electrical power consumption because of recent utility rate increases and adoption of MOS (metal oxide semiconductor) technology.

The 1974 budget survey also indicates that workload in terms of jobs per month or in total transactions increased 20%. The net effect is that the data processing manager is now under pressure to produce more output for very little more money. [McLaughlin, 1974 and 1975]

II. NATURE OF THE PROBLEM

Management control of the computer resource, like that of any staff activity, is concerned with answering the following:

How much money should be spent on it?

How should the money be deployed for maximum effectiveness?

Is the resource being employed efficiently?

This thesis will present an analysis of alternative systems and techniques for controlling the computer resource with primary emphasis on the allocation of services of an organization's existing data processing facility.

Although the goal of the computer is a straightforward economic one--to help operating units and staff personnel to execute their responsibilities better through less costly processing of data, improved organization of information and procurement and deployment of information that is too expensive to obtain otherwise--the resource has a complex set of supply and demand characterisitcs and is still relatively new. [Dearden and Nolan, 1973]

Some companies have found that their computer budgets have increased over the years, quite independent of their overall growth. In many organizations, there is simply no attempt made to measure the benefits or lack of benefits that are resulting from data processing expenditures. Reliable studies over the past several years have found that, with a few notable expections, most companies have a very poor idea

of the benefits that are realized from their use of computers and where they should look for further applications. A far more serious matter, however, is when companies make no attempt to measure the cost aspect.

Management must continue to emphasize efficiency; which is the conversion of inputs into outputs with as little loss as possible. However, increased importance must be placed on effectiveness; performing activities which are congruent with the overall organizational goals and objectives.

Too often the data processing professional or his customers are more interested in pursuing their own objectives than those of the organization. McLean has described this misdirection of attention as "committing a Type III Error," or solving the wrong problem. [McLean, 1973]

III. ECONOMIC THEORY

Application of economic theory alone will not solve the allocation problem but it does provide a frame-work for analysis.

A. ECONOMICS OF COMPUTERS

There is general agreement concerning the set of supply and demand characteristics of computers. [Smidt, June and Fall 1968, Nielsen, 1968 and Dearden and Nolan, 1973]

1. Supply Characteristics of Computers

There is a high ratio of fixed to variable costs. Hardware and systems development costs are high, but variable operating costs are low. Thus cost per unit of computing is greatly influenced by total usage of the data processing facility. This provides an incentive to maintain a full-capacity workload since the cost of incremental work is so very low. This high ratio leads to difficult priority problems once the computer system is operating at full capacity. Determination of which new demands should replace existing demands can be very difficult.

There are significant economies of scale in computer equipment. Particularly with the central processing unit, an expansion in capacity can result in a less-than-proportional increase in costs. Thus it is often logical for all organizational units to share the use of a common computer facility.

Because computers are subject to significant indivisibilities, incremental capacity must be acquired in large

blocks. Therefore, a linear demand growth, which is not necessarily always the case, cannot be accommodated smoothly.

2. Demand Characteristics of Computers

The demand for computer services has been growing rapidly in complexity and size. The initial thrust was to computerize manual operations. This has been followed by more sophisticated on-line applications involving data base management and management information systems. Because technological advances have consistently reduced the marginal cost of computing there has been increasing pressure to use computers in labor-intensive tasks.

There are daily, weekly, monthly, annual and close-of-business cycles in the demand for computer services. Smoothing these peaks and valleys presents formidable management problems.

Projects and programs are postponable depending on the user and the timing. Consistent handling of priority variations, reflecting user costs and inconvenience, is not an easy task.

A single computer system will not provide optimal services for all business and scientific applications. System configuration must be balanced between the requirements of business applications which tend to be input/output bound and scientific applications which require large high-speed computational capability. This may result in some suboptimal computing.

B. MICROECONOMICS AND RESOURCE ALLOCATION

Discussion of this resource allocation problem requires consideration of classical microeconomic theory as it relates to an organization or firm having a computer service center whose primary customers or users are units of the same organization or firm. It is generally concluded that the optimum point for the organization is where the marginal revenue of an additional unit of work is equal to the long run marginal cost. [Smidt, Jun 1968, Sharpe, 1969 and Subzak, 1974] Several assumptions are made to simplify this analysis and discussion of Figures 1 through 7 which are contained in Appendix A:

Computer customers will adjust demand in a cost sensitive manner.

Total value of service will increase at a decreasing rate as a result of the law of diminishing returns.

Capacity is infinitely divisible (as indicated earlier, this is not the case with computers but this assumption is necessary to simplify the theoretical analysis).

In Figure 1, the TC (total cost) curve represents the cost to the computer center, and therefore to the organization, of providing capacity for a specific quantity of computer work (Q is used to represent units of computer work). The TV (total value) curve represents the value to the user, and therefore to the organization, of that work. A total cost of TC_1 dollars is required to produce OQ_1 work units which have a value of TV_1. TV_1 minus TC_1 equals net benefit to the organization.

Figure 2 represents the MC (marginal cost) and MV (marginal value) curves that correspond to Figure 1 total curves. The optimum point is to produce Q_1 units at price P_1. Producing additional units will result in spending whose value is less than the cost. Producing less units will result in unsatisfied demand which could be fulfilled by less-than-proportional spending.

As indicated earlier, computers afford economies of scale. Therefore, the LRMC (long run marginal cost) curve for computer services is downward sloping when the organization is operating within the long run economies of scale segment of a "standard" cost curve. To be optimal, the amount of computer capacity must simultaneously satisfy two conditions:

The total costs of a capacity level must be less than or equal to total value.

The marginal cost must equal the marginal value.

In Figure 3, both conditions are satisfied if capacity Q_1 is available at price P_1. Condition one is satisfied since the total value, area $ODBQ_1$, exceeds the total costs, area $OABQ_1$.

To see that condition two is also satisfied, we shall expand Figure 3 as shown in Figure 4. Suppose that initially OQ_2 capacity was available and consider the incremental costs and values of providing an additional Q_2Q_1 unit of capacity. This increase in capacity would increase total costs by an amount equal to the area Q_2FBQ_1. The excess of value over costs are given by the area EFB.

If the optimal capacity was supplied, the optimal allocation of this capacity would require an average price Q_1B. The total revenue received would be shown by the area OQ_1BP_1. This area is less than the total costs, area $OABQ_1$ of the optimal capacity. This analysis leads to the conclusion that the computer resource would be a deficit operation and would not necessarily provide for full cost recovery under these conditions. Since computing equipment is not infinitely divisible, optimality requires that the smallest feasible increment in capacity be provided as soon as the incremental value of this capacity equals its incremental costs. [Smidt, Jun 1968]

Classical microeconomic theory can also be used to show that if a computer center attempts to operate as a profit center, suboptimal conditions will result.

In Figure 1, the total revenue to the computer center corresponds to the TC curve. Operation as a profit center requires the addition of a revenue curve above the cost curve. The TV curve in Figure 5 represents the total value to the customer and also to the organization. The TR curve, however, represents the cost curve to the customer division and the TV curve to the computer center as the objective now is to maximize computer center profit. The TR curve is below the TV curve since the customer will not pay a price higher than the value received. The profit maximization point for the computer center is now where the marginal revenue equals

the marginal cost. Figure 6 shows this quantity at Q_2 and the price P_2 which is still less than the value to the customer.

Figure 7 shows the total profit to the organization is equal to TV_2 minus TC_2. This profit is divided between the computer center and its customers with the center receiving a profit equal to TR_2 minus TC_2 and the customers' profit equal to TV_2 minus TR_2. However, the total profit to the organization, TV_2 minus TC_2, is less than the original total profit, TV_1 minus TC_1, as shown in Figure 1 when the computer center is operated as a cost center. Therefore, the theoretical maximum profit contribution to the organization cannot be greater than the profit realized when the computer center operated as a cost center. [Sharpe, 1969 and Sobczak, 1974]

C. PRICING AND RESOURCE ALLOCATION

The economic system of any society must solve the resource allocation problem. In the United States, the price system is the dominant allocation technique. Like other allocative mechanisms, the price system does not always work as well in the real markets as in theory. Therefore, non-price allocation techniques are used for certain categories of goods and services at certain times to attain social objectives. Governmental controls over production and distribution during wartime are classic examples.

Most resources can be used to produce many goods which are useful to many consumers. Due to the scarcity of resources, those used by one producer are not available to any other and goods consumed by one person reduce the total consumption

possibilities of others. Thus, the preferences of different economic units for the same economic resources must be determined.

In our market economy, the price system is used to express the preferences of economic units concerning what to produce, how and for whom. Theoretically, prices will be bid up or down to the point where the available supply is allocated to those willing to pay the highest prices. It is assumed that the value of the goods to society is the maximum price paid for the goods by any consumer. A price system can dynamically allocate resources by communicating information concerning consumers' demands to producers and providing consumers information about the cost of satisfying different wants.

To equate supply and demand, prices must be allowed to respond in certain ways. When demand rises or falls, they must rise or fall sufficiently to ration the available supply. Over time, changes in supply must be allowed to affect the price of any good. If the demand for a good is low, its price must be permitted to fall, even below cost, so the proper information about demand may be obtained and so the allocation of resources can adjust properly to the unprofitability of that good. Therefore, at any instant in time prices need not bear any precise relation to the cost of production at that time.

In summary, prices are not a cost recovery mechanism but are a device for allocating scarce resources and obtaining the efficient level of investment over time. [Singer, Kanter and Moore, Fall 1968]

D. PRICING AND COMPUTER RESOURCE ALLOCATION

Whether or not a computer service facility operates with monopoly power, e.g., customers are not permitted to purchase data processing services from sources outside the organization, or operates as a market competitor should be covered by organization policy. Pricing is a feasible allocative device in both situations when certain conditions are satisfied.

Computer time should be made available to any customer willing to pay the price. A budget constraint, usually in the form of dollars or time, must be imposed on the customer to provide an incentive to evaluate the benefits of the computer service relative to its costs to him and other customers.

A user of the service facility imposes costs on all other users by increasing their turnaround time. Therefore, the price each customer pays should be high enough, in principle, to compensate all other customers for any reduction in service. If customers are allowed to bid up prices, equilibrium will be reached where each customer is paying a price which is greater than the total value of the diminution in service to all other customers.

A final condition for effective pricing is that prices must be free to flucturate in relation to the forces of demand and supply, not the cost of production. [Singer, Kanter and Moore, Fall 1968]

IV. ACCOUNTING THEORY

From the financial control point of view, the primary function of the data processing accounting process is to maximize the benefits the organization obtains from the data processing resource by providing managers with the information required to make prudent decisions. While this goal is generally accepted, there is a variety of procedures in use for the accumulation and presentation of this information.

A. BUDGETING

Budgeting systems are designed to accomplish a multiplicity of functions: planning, coordinating activities, implementing plans, evaluating performance, communicating, motivating and authorizing actions. In governmental or not-for-profit budgeting, budget appropriations primarily serve as authorizations and ceilings for management actions.

The budgeting system must be designed to encourage managers to act in harmony with the overall objectives of the organization. When administered wisely, budgets will compel management planning, provide definite expectations that are the best framework for assessing subsequent performance and promote communication and coordination among the various segments of the organization.

Although the human factors in budgeting are considered more important than the accounting procedures, budgeting is too often looked upon from a purely mechanistic viewpoint. The success of a budgetary system depends upon its acceptance

by the persons who are affected by the budgets. Ideally, attitudes are sympathetic, cooperative and cost-conscious. [Horngren, 1972]

B. SERVICE CENTER COSTING

There are three interdependent aspects of cost allocation: (1) choosing the cost object, (2) choosing and accumulating the costs that relate to the cost object and (3) choosing an allocation base to associate the accumulated costs with the cost object.

1. Marginal Versus Full Costing

There is no general agreement regarding the proper approach to costing. Some accountants support marginal costing on the theory that if customers paid less than marginal cost they will be motivated to use more service than economically justified. Also, if customers paid more than marginal costs, they may choose to not use the service because the value to them is not equal to marginal cost. This would be a mistake if the value of the service to the organization was more than the marginal cost.

Others support full costing (marginal plus a share of fixed costs) of services on the premise that something is wrong with service quality or efficiency when services are not worth their full cost. Also, the service center may be favorably motivated to keep allocated costs down in order to avoid customer complaints.

Whether to use marginal or full costing must be ultimately resolved in terms of how the alternatives influence

management behavior in a particular organization. If the use of the organization's service center is not controllable by the customer, then it may make little difference whether or not fixed costs are allocated. A plausible solution is that:

> If management does not want the service center to compete with outside sources, use marginal costing.
>
> If management desires the service center to compete with outside sources, use full costing.

2. Principles of Cost Allocation

There is general agreement that if certain costs are to be allocated, customer charges should not depend on another customer's level of consumption and that only budgeted costs should be allocated. The latter principle is based on the opinion that the service center not the customer should be held responsible for actual price or efficiency variances within the service center. [Anthony, Dearden and Vancil, 1972 and Horngren, 1972]

V. ALTERNATIVE SOLUTIONS TO THE ALLOCATION PROBLEM

The information presented thus far should be briefly reviewed before considering alternative solutions to the problem of allocating the services of an organization's existing data processing facility:

The data processing resource represents a significant dollar investment which has increased 8%-13% annually in the recent past. Expenditures are significantly affected by inflation and material shortages.

Management control is becoming increasingly difficult because of advances in hardware and software technology and the rapid growth in the complexity and size of the demands for computer services. Increased importance must be placed on performing activities which are congruent with the overall organizational goals and objectives.

Microeconomic theory supports the conclusion that the theoretical maximum profit contribution to the organization cannot be greater than the profit realized when the computer center is operated as a cost center. Also, that when optimal capacity is supplied--total costs are less than or equal to total value and marginal cost equals marginal value--the computer resource will be a deficit operation and would not necessarily provide for full cost recovery.

If prices are used as a resource allocation mechanism, they must be allowed to fluctuate in response to supply and demand. They should not also be considered a cost recovery

mechanism in the short run since at any instant in time, prices need not bear any precise relation to the cost of production at that time.

Budgeting and accounting systems should provide managers with the information required to make prudent decisions. They should also encourage managers to make cost-benefit trade-offs in a manner which is beneficial to the entire organization. If certain costs are allocated, customer charges should not depend on another customer's level of consumption and only budgeted costs should be allocated.

Solutions to the data processing resource allocation problem can be sorted into two main categories:

No charge-back - computer center costs are treated by the organization as an overhead item and if distributed, are handled in the same manner as other overhead expenses.

Charge-back - customers are charged for their use of the data processing center in proportion to their use of the center's services.

A. NO CHARGE-BACK APPROACH

When the data processing resource is treated as an overhead item, users are encouraged to increase computer usage. They improperly view the data processing resource as a "free good" when in fact, the dollars invested in providing the resource are scarce goods to the organization. Users will prefer to substitute the "free good" for other resources for which they must directly pay.

Although users can be given accurate service cost data, on a continuing basis under a no charge-back system, they have no incentive to make efficient and effective use of the data processing resource. The computer center may be misled into overestimating computer demand, resulting in overinvestment in computer facilities.

Since users are not directly accountable for the cost of employing computer resources, some form of centralized administrative control is required. This is typically accomplished by a steering committee composed of computer center management and user representatives. Control of the committee and the data processing resource is usually gained by computer center management because that organizational unit is the only one which has a financial commitment to and the technical knowledge in the data processing function. The control process becomes biased by the dependence on advice from computer center management which, in effect, is helping decide what financial support should be given its own function.

Another common administrative technique is to establish two types of priority rules: one which determines the access pattern of a given set of users and another which places a valuation on the importance of user projects.

The FCFS (first come first served) and the SJF (short job first) procedures are examples of the first type of priority rule. The difficulty with procedures of this sort is that an implicit assumption must be made about the value placed on computer time by each user or the cost of delay to each user.

In general, users will not value time equally nor consider waiting equally costly. Therefore, such procedures will not allocate time properly.

A FCFS allocation procedure is not necessarily a bad one, if in fact, all projects are of equal importance. If this is not the case and users are permitted to submit as many projects as frequently as they desire, delays in turnaround will act as a rationing system. As delays grow longer, demand gradually becomes choked off because users eventually reach the point where nothing can be done until previous projects are returned.

The SJF procedure is imposed to reduce average turnaround time. However, undesirable results may be experienced if the loss from delayed receipt of a long project greatly exceeds that associated with short projects.

The second type of priority rule is one which rations access to the computer according to the relative importance attached to the user's project. These priorities may be viewed as a form of artificial money since a high priority has the same effect as a customer with a large budget for buying data processing services. Unless priorities are reviewed and reset periodically to reflect user importance, some projects are likely to have either too easy or too hard a time gaining access to the computer resource. Also, it is questionable that all users within a project should receive the priority attached to that project.

A priority system for access to the computer resource discourages efficient substitution of other resources for computer

use. Functions which should not be computerized will be by the high priority user just because his project priority is high.

The heart of the problem with centralized administrative control through a steering committee or various administrative regulations and exemptions is that allocation decisions will be made by the individuals least qualified to make these types of decisions. Under the no charge-back approach, data processing center personnel, including computer operators, may be permitted to make allocation decisions based upon friendships or compassion for an individual who can tell a convincing story. While these individuals may be technically qualified in the data processing field, they probably are not qualified to judge the value of various projects to the organization. Therefore, it is concluded that there must be better means for resolving the data processing resource allocation problem. [Singer, Kanter and Moore, Fall 1968, Sharpe, 1969, Nielsen, 1970, Dearden and Nolan, 1973 and EDP Analyzer, 1973 and 1974]

B. CHARGE BACK APPROACH

Implementation of a charge-back system implies that customers will include the prices established for data processing services as a factor in the decision-making process and react rationally to those prices. It is believed that since prices are used to allocate the limited resources of our national economy, it is reasonable to expect that pricing can also be used to allocate data processing resources.

A relatively simple test of the effect pricing can have on data processing customers was conducted at the Campus Facility of the Stanford Computation Center a few years ago. Customers were reserving magnetic tapes at an ever increasing rate; yet they were releasing very few of them. Pleas were made for customers to release all tapes that were not absolutely essential to their work, but with little success. Computer center management then decided to levy a nominal charge of $1 per tape per month. On the first day that the charge was instituted, customers released more than one-third of the "absolutely essential" reserved tapes. [Nielsen, Fall 1968] This test indicates that pricing can be used to motivate customers and to attain organizational objectives.

Most of the literature cited in the Bibliography discusses some of the arguments supporting charge-back systems. While the following arguments favor charge-backs they do not necessarily favor cost recovery versus resource allocation.

A charge-back system gives the customer some control over both the cost and the quality of data processing services. There is an economic motivation to make trade-offs between value and cost resulting in more effective use of available resources.

With an adequate pricing structure, a charge-back system will do a good job of resource allocation. Although charging for data processing services involves an administrative overhead expense in manpower and computer time, it can lead to the more effective use of scarce resources which will more

than compensate for the cost of the charging system. The computer center may not perform additional processing under such a scheme, but the work it does perform is more valuable to the organization.

A well designed charge-back system encourages better resource utilization by providing an incentive for customers to buy off-peak hours.

Charge-back systems permit decentralization so that various levels of management are given the authority, opportunity and responsibility to make decisions for which they are best qualified. This improves the quality of decision-making for the good of the organization.

1. Pricing Goals

Charge-back systems can be utilized to simply recover the costs of the data processing resource in some manner related to actual use. When the data processing facility is organized as a profit center, the charge-back system goal is not only to recover costs, but also to provide profits. In both cases, charges will be closely related to the costs of providing those services.

Another goal of charge-back systems might be to provide effective economic allocation of the data processing resource. That is, higher prices are charged for scarcer resources. Since the price structure is designed to influence customer behavior in the manner desired by management, charges used to attain the goal of effective economic allocation may have little relation to the costs of providing the services.

Another plausible goal would be the combination of cost recovery and resource allocation. Prices are set not only to allocate the available resources in the desired manner, but also so that the usage of the resources will provide sufficient aggregate revenue in the long run to recover the costs of the data processing resource. [Nielsen, 1968 and EDP Analyzer, 1973]

2. Pricing Principles

Various pricing techniques, average cost, flat rate and flexible pricing will be covered in detail in Section V. B. 4. Regardless of the established goal(s) and the pricing technique employed in a charge-back system, there is a body of pricing principles upon which the pricing system should be based.

The first of these is the principle of resource utilization. The price schedule must properly motivate the customer and generate sufficient revenue to meet the established goals.

The second principle is that of price reproducibility. Customers desire stable charges in order to better control expenses. A project should cost about the same amount each time it is processed assuming the same level of service.

The third principle is that of customer control. A differentiation must be made between the use of resources over which the customer has control and those which he cannot influence.

A fourth principle is that of customer understandability. It requires that not too many factors be included in the price schedule and the data processing accounting algorithm. If the pricing procedure requires too much of the customer's own time or extensive training, he may ignore the pricing system altogether.

An obvious but easily overlooked matter is that the data processing center should not develop the pricing system by itself. An effective pricing system requires the participation of many other people in the organization, such as financial and planning personnel and customers. [Nielsen, 1968, Hootman, 1969, and EDP Analyzer, 1974]

3. Pricing Structure

Several characteristics of the pricing structure which warrant close evaluation in terms of the organization's particular data processing system are: length of costing period, selecting the resources to be priced and the quality of service.

a. Length of Costing Period

The first aspect of the length of the costing period is concerned with the pricing interval. It is the shortest interval of time during which no change in the assigned price for computer service can occur. Short pricing intervals and thin demand are likely to produce highly erratic prices which in turn will add to the difficulty faced by the customer in estimating costs. Increasing the length of the pricing interval will reduce price variation but, if the

pricing interval is too long, customers requesting high priority service will tend to be charged the same price as customers requesting low priority service. This reduces the incentive to economize on use of the data processing resource.

As would be expected, no one set of pricing intervals is proper for all organizations. It does appear desirable, however, to have multiple pricing intervals or shift differentials within a twenty-four hour period. When demand is low the pricing interval should be longer than the interval when demand is high. The actual number of intervals may also depend on shift length and the number of operating shifts. [Smidt, Jun 1968]

The second aspect of the costing period is how frequently the basic price schedule is revised. Customers desire stable prices over considerable periods of time. The data processing center, however, may very well want more frequent price adjustments to improve the likelihood of achieving its cost recovery goals.

The customer has a fair degree of flexibility in selecting the resources to use for a particular project during the original design or over time as his project is modified. However, he usually does not have the flexibility to readily change the mix of resources required within a very short period of time. Recommendations in the literature [Nielsen, 1968 and EDP Analyzer, 1974] range from monthly to annual revisions of the price schedule. Therefore, it is recommended that each organization adjust the life of its price schedule to

meet the capabilities of its customers to adjust the resource requirements of their projects.

b. Selecting the Resources to be Priced

This is the most complex facet of the pricing structure. It can become an allocation problem itself since a charge-back system is based on pricing resource usage and a single project will normally utilize several resources which have alternative units of measure. Printing alone can be measured in minutes, lines or pages. Another difficulty involves the concept of demurrage. This is charging a customer for a resource he is not using directly because the resource is not available to any other customer.

Selection of the proper resources and utilization measures involves trade-offs among the following:

The utilization measure must be easily obtained from the host system. It is not desirable to spend a large portion of the system's resources just trying to account for utilization.

Increasing the number of prices in the system can improve resource allocation. However, this must be compared with the time the customer will utilize considering price schedule options.

Assigning resource measures only to those resources over which the customer has control encourages good decision making. But, are these resources the more expensive or critical resources to the organization? [Nielsen, 1968 and 1970]

A simplified approach to this complex problem is to develop measures of work which in turn are priced, e.g., the amount of work that can be performed on a computer system having an annual rental of one million dollars. These are called CRU's (computer resource units) by Bell Telephone Laboratories and MU's (machine units) by the National Institute of Health. These similar but not identical approaches combine several resource measures into one measure of work.

Customers are provided with a simple bill for so many CRU's or MU's at a stated price. Such a system can provide for cost recovery, but it can also defeat the goal of efficient resource allocation by hiding the prices of individual resources. This is especially true when two different types of equipment perform the same function but with significantly different capabilities. Also, it is impractical for a CRU or a MU to remain stable as equipment is changed. [EDP Analyzer, 1973]

The CAMP (cost analysis measurement plan) utilized by the Pacific Telephone Corporation is a composite cost recovery system but not to the extent of the single computer center rate under the CRU and MU systems. It recognizes that all projects do not use the same proportion of computer system resources and that it would be unfair and inaccurate to establish a single rate and to charge this rate regardless of the kind or mix of resources used.

The computer system is divided into cost pools which are groupings of machine related costs according to a

common statistic or rate. Customers are charged one rate for the use of any resource in the pool. The three basic types of cost pools are: peripheral, main memory and processor. Separate cost pools are recommended for each type of peripheral device. Various methods are also utilized for allocation of some direct and indirect costs.

The CAMP system effectively transfers the cost of data processing services from the computer center to the customer. Also, it has favorably modified his tendency to consume data processing resources without discrimination as was experienced when data processing services were available at no cost to the user.

The CAMP system, however, has failed to reap the full benefits of more rational resource allocation decisions for two reasons:

The customers do not understand the system. The CAMP Manual is so large and complex that most management personnel may not have even attempted to read it.

The price schedule does not provide sufficient flexibility to encourage cost-benefit decisions beyond the determination of whether or not to mechanize a particular function. The allocation of personnel costs based on the actual level of machine usage by the organization places this significant cost factor beyond the control of the individual customer. Also, the direct machine usage price is determined by dividing the monthly rental by the possible number of hours the device can be used. This may be considered uneconomic pricing in that all customers pay the

same rate whenever they choose to utilize the resource. No adjustment is made to reflect the quality of service or the temporal aspects of resource usage. [Mobley, 1974]

Review of the composite or pool pricing structures indicates that in order to allocate the entire data processing resource so as to maximize its economic use, it is necessary to have an array of prices upon which customers can base cost-benefit decisions. A list of sixteen potential utilization measures and eighteen possible relations between measures and resources is shown in Tables I and II of Appendix B. They provide a basis for development of a price schedule for resource usage in terms of utilization measures which are understandable and controllable by the customer. It should be recognized that software design or hardware availability may preclude utilization of some of these measures in any particular data processing center. [Nielsen, 1968]

In order to indicate the kinds of consideration behind these utilization measures, six input-output measures will be discussed briefly. Input-output time and number provide a measure of a project's usage of various peripheral devices, control units, channels and peripheral processing units. Measurement in terms of number of cards and lines provides an indication of productive work which is more likely to be meaningful to the customer than the number of seconds of card reader time on a certain reader. This is particularly true when there are similar devices having different speeds and when the customer has little or no control over

the actual device employed. These measures also eliminate the customer's concern about card jams, paper breaks and other incidents over which the customer has little control.

A portion of the personnel costs are covered by a measure of set ups which is provided to indicate the special work which is performed by computer center personnel to enable a project to be processed. This would include mounting of reserved tapes, the mounting of private disk packs and the loading of special forms on a printer. A more comprehensive explanation of the tables in included in Appendix B.

Under many older batch processing systems, a customer can tie up all of the resources of the computer center even though he is not actually using all of them. Under this condition, it may be quite satisfactory to charge the customer a single price.

In today's multi-programming environment, there is no single resource that can be priced in order to allocate the usage of the entire system. Also, a single customer need not utilize all of the resources of the computer center at any one time since it is possible that other customers may be able to utilize some of the remaining resources. [Nielsen, 1968]

The main problems of developing an equitable charge-back system in a multi-programming environment involves charging for shared and allocated resources. Shared resources are those used by a project which are made available to it only for a portion of the project's elapsed time. Allocated

resources are those dedicated to a project for the full elapsed time of that project or project step.

Elapsed time or wall clock time is not equitable for those resources because of involuntary wait time which can be large and is uncontrollable by the customer. A possible solution is to subtract recorded or estimated involuntary wait time from actual elapsed time to determine productive time for charging. In the present state of the art, this approximation method appears to be the best technique for arriving at customer charges in the multiprogramming environment. [EDP Analyzer, 1974]

The pricing of specialized hardware also presents an interesting pricing problem. A customer might require some specialized service for which it might not be feasible for the computer center to recover its costs through the regular charge-back system.

A proposed solution is for the customer to purchase the equipment and reimburse the center for the incremental operating costs. If another customer used the equipment, they would be charged a "reasonable" price and the charges would be credited to the owning customer. [Smidt, Jun 1968]

c. Quality of Service

Turnaround time is the most popular measurement of service quality. A more complete description is - desired project completion time and possible variance in that completion time. There are significant variations of project completion times. Some may require immediate response.

Others may need completion by "Friday." Still others may desire completion as-soon-as-possible which can have a wide range of customer values - anytime this morning, anytime today or anytime this week. [Sharpe, 1969]

The heavier the demand on the computer resource, the longer become queues, completion times and the variance of completion times. Streeter's analysis of a scientific computing environment indicates that response time grows rapidly when a data processing system is utilized above 70%. It also pointed out, that the relative value of the computer center to its customers decreases as response time increases. The study concluded that the total relative value of the computer center reached a maximum when the data processing system is loaded to just under 80% of capacity and the average response time was just under two hours. [EDP Analyzer, 1973]

There is clearly a difference in the value associated with different levels of service quality yet, there may be no cost differential in providing them. Economic pricing therefore, requires recognition of the temporal allocation of the data processing resource which involves the time when the physical resources are provided as well as recognition of its physical measurement and allocation. [Nielsen, 1970]

Computer time is not a homogeneous commodity. Time during the day is usually preferred to computer time at night. Utilization of a single price for all service levels can reduce the value of the computing service to certain

customers and can encourage overinvestment in hardware. If the price of two and twenty-four hour turnaround times is the same, customers will usually select the more convenient two hour service. This can adversely affect the turnaround time to customers which require prompt service without increasing the value of service to the customer whose actual need can be satisfied by twenty-four hour service. Without proper validation of service level requirements, the organization may install additional hardware to meet inflated peak workloads. [Kanter, Moore and Singer, Jul 1968 and Sharpe, 1969]

A price schedule which reflects the quality of service provided encourages customers to make cost-benefit trade-offs. An approach which maximizes the value to the organization of the computing actually performed is as follows:

> Raise the price for shorter turnaround service.
>
> Customers can then select a level of service and pay whatever is required to obtain it.
>
> Customers can also select a price level and accept whatever level of service that can be obtained at that price.
>
> [Smidt, Jun 1968 and Nielsen, 1970]

In summary, when selecting the resource to be priced, it is desirable to select as few measures as possible, but these must be measures of service over which the customer has substantial control and which taken together account for a major portion of the system's resources.

4. Pricing Techniques

Any charge-back system, whether designed to provide for cost recovery alone or cost recovery and resource allocation, should consider the pricing principles and structures discussed in the previous sections. Most pricing techniques provide for cost recovery in the long run. "Data processing costs are incurred for the customers' benefit so why not charge them for this service?" or so the adage goes. This section will discuss the average cost and flat rate techniques which primarily provide for cost recovery and the flexible pricing technique which also provides for economic allocation of resources. Pricing techniques for operating the data processing resource as a profit center will not be included. Earlier discussion of microeconomic theory indicated that operating the data processing facility as a profit center could not improve the theoretical maximum profit contribution to the organization above that realized when it is operated as a cost center.

a. Average Cost

Under the average cost technique, total actual data processing resource costs are charged to the customers. The basis for charging is the actual hours used in servicing each customer. The charging rate is determined by dividing the total costs by the total number of hours of the data processing resource which customers employed.

A net decrease in data processing resource usage results in increased hourly charges. Customers are then

motivated to decrease their usage. This in turn, results in further rate increases for the data processing service. Note also, that such a cost allocation system results in perverse resource allocation when usage increases. Rates are driven down attracting more usage, and so on to the point that when facility capacity is reached, price is at its minimum, and customers can afford to use the computer resource for trivial functions.

Additional faults of most cost allocation systems can be illustrated by referring to Figure 8. Assume that an organization consists of two operating departments (customers) and one service department (computer center). The computer center's total annual costs are being allocated to each customer on the basis of machine hours utilized. The computer center's monthly costs consist of $6,000 in fixed costs plus variable costs of $.40 per machine hour.

The top section of Figure 8 shows customer allocations at a 10,000 hour service level. The lower section shows the results when one customer's usage decreases by 2,000 hours. The other customer's annual cost increases 15% even though its usage of the data processing resource remained constant. [Horngren, 1972] Figure 8 is on page 42.

The average cost technique has the following significant deficiencies:

The charging rate depends on the level of utilization by the organization, not the individual customer. Computer operations have a high fixed cost and a low marginal

FIGURE 8

Average Costing of Service-Department Costs

At a 10,000 machine-hour level of the production departments

Actual costs = \$6,000 + \$.40 (10,000 hours)	=	\$ 10,000
Rate per hour = \$10,000 ÷ 10,000 = \$1.00		
To Department 1: 5,000 hours X \$1.00	=	\$ 5,000
To Department 2: 5,000 hours X \$1.00	=	5,000
Total reallocated		\$ 10,000

At an 8,000 machine-hour level

Actual costs = \$6,000 + \$.40 (8,000 hours)	=	\$ 9,200
Rate per hour = \$9,200 ÷ 8,000 hours = \$1.15		
To Department 1: 5,000 hours X \$1.15	=	\$ 5,750
To Department 2: 3,000 hours X \$1.15	=	3,450
Total reallocated		\$ 9,200

cost. Therefore, when utilization is high charges will be low. This will attract additional demand on an already overloaded computing resource which is just the opposite of what might be desired. Excessive system expansion is encouraged because demand is high and all costs will always be recovered. If additional capacity is added, charges will suddenly jump upward because of excess capacity. Then the price increase will discourage utilization of the newly added capacity.

The charging rate is very sensitive to the length of the costing period. Emphasis must be placed on the long run matching of cost and charging revenue. Monthly price schedule revisions would normally result in highly erratic rates which are undesirable to the customer.

Average costing completely delegates decisions on use of the data processing resource to the customer. It is rare that a group of independent decisions made by the customer will optimize the organization's overall use of the resources. Since prices do not reflect the quality of service, bottlenecks may result around critically scarce resources within the system. Customers may use them wastefully since the penalty for being wasteful is small.

The data processing facility does not have an incentive to be efficient. All costs resulting from computer center inefficiency are borne by the customers.

The incentive to improve the efficiency of projects is minimized because the elements in the pricing system are practically invisible to the customer.

This charging technique does not facilitate the cost versus quality of service trade-offs. The concept of value of service is not reflected in a price schedule. With no price differential between popular and unpopular hours, customers will select the popular hours.

Because of these deficiencies, various administrative controls are usually placed on resource usage. The shortcomings of administrative procedures were previously discussed in Section V. A.

A variation of the average cost technique is to apply a shift differential to the basic rate. This modification provides some incentive for the customer to make cost-benefit trade-offs. The University of California Computing

Center uses three priority levels--urgent, standard and delayed. Projects assigned the urgent priority are charged 130% of the basic rate (except for resources not affected by priority, such as the cost of paper). The delayed priority is charged at 80% of the basic rate. Center customers have revised their projects in accordance with the price structure and the turn-around constraints.

Average cost is currently the most popular pricing technique. The Federal Government requires use of the average full cost technique for identifiable direct costs or the indirect costing technique on cost reimbursable contracts. Use of average cost is encouraged. This inhibits utilization of excess capacity by unfunded customers since increased usage would reduce the average cost per unit of usage and thus the amount of Federal Government reimbursement. Under average costing, it is relatively simple to record usage and bill customers accordingly. However, it does a very poor job of resource allocation. [Smidt, Jun 1968, Kanter, Moore and Singer, Jul 1968, Nielsen, 1970, Horngren, 1972, Dearden and Nolan, 1973, and <u>EDP Analyzer</u>, 1973 and 1974]

b. Flat Rate

The flat rate technique is very similar to the average cost approach. The basic rate, however, is calculated in advance based on anticipated available capacity and estimated effective usage for the forthcoming pricing interval. Therefore, the resulting revenue to the data processing center may be somewhat less or greater than the costs.

The customer is provided with a basic rate before he "buys" any services, but he is not much better off since cost is nearly independent of turnaround or service. The deficiencies enumerated earlier for the average cost technique also apply to the flat rate technique. [Nielsen, 1968 and EDP Analyzer, 1973 and 1974]

c. Flexible Pricing

Flexible pricing is another popular charge-back technique. It provides for cost recovery in the long run and influences how the data processing resource is utilized. Since prices are established to match supply and demand, the resource is priced at its economic value which has little or no relation to cost at any point in time. The price schedule fluctuates in such a way that prices are high when demand is large relative to capacity and low when demand is small relative to capacity. Its price structure includes identifying the critical resources, management's desires to influence the workload and the political realities of the organization.

Peak-load pricing is a similar technique which many industries utilize when the demand for their service or product varies greatly at different times in relation to the capacity available to meet that demand. Airlines charge more for midday flights than the midnight coach. Telephone companies vary their rates by time of day and by time of the week to encourage more efficient use of plant and equipment. The Wall Street Journal recently reported that some electric

utility companies were experimenting with peak-load pricing. Most utilities experience sharp peaks in demand during mid-morning and the late afternoon. There are also seasonal peaks in different parts of the United States. Although the average load factor is 53%, utilities must invest in substantial additional capacity to meet peak demands. Difficulties in actual implementation involve measuring consumption and obtaining the approval of customers and rate setting commissions. [Wall Street Journal, 1975]

If flexible pricing is utilized to allocate the data processing resource, the price schedule should influence customer behavior regarding resource usage. Certain factors which may be included are:

An administrative regulation prohibiting long projects during certain time frames may not be required when the central processing unit price is established as an increasing function of compute time, e.g., the price for the second five minutes is twice the price for the first five minutes, etc.

To encourage running long projects at night, use a decreasing function of compute time to determine the actual charge for that shift.

If the organization has a hierarchy of on-line file storage devices, there can be a hierarchy of storage prices with each price being related to speed, access time and transfer rate of a class of devices. This structure allows the customer to make cost-benefit trade-offs which cannot be done under cost pooling.

Since short projects are relatively insensitive to usage price differences, fixed charges may be added to reflect different levels of service.

Fixed charges can also be included in the price schedule to discourage certain types of system utilization. The casual punching of cards can be discouraged by levying a fixed charge in addition to a flat price per card punched. This approach makes a few cards relatively expensive but would have little effect on the total charge for large punching projects.

The extent to which the above behavioral factors can be included in the price schedule depends on the configuration of the data processing system, the capabilities of its operating system and careful evaluation of the trade-off between improved resource allocation and maintaining the price schedule at a level which is understandable and meaningful to the customer. [Smidt, Jun 1968, Nielsen, 1968, EDP Analyzer, 1973, Dearden and Nolan, 1973 and Bray, 1975]

There is a large variety of queuing disciplines which may be reflected in the price schedule in an attempt to minimize the sum of the costs of waiting and the costs of providing the service facilities. Choice depends upon customer preferences concerning uncertainty in regards to turnaround time and/or a simple price schedule versus a complicated one. Examples of possiblities are:

Different prices per minute on different computers. The lower priced machine would be expected to have a heavier load and longer turnaround times.

Provide several queues with different prices on the same system. The higher the queue's priority, the higher the price. A queue is serviced only when all higher priority queues are empty. Also, serve on a first come first served basis within a queue. [Bierman, Bonini and Hausman, 1969, and Sharpe, 1969]

Other theoretical proposals involve bidding or quotation systems using highly volatile pricing of short term requirements. Under the bidding system, the customer actually "bids" against other customers for the use of scarce resources. With a quotation system, a quotation program keeps track of supply and demand and adjusts prices accordingly. The customer requests a price quote for a particular quantity and level of service. After considering the reply, the customer can accept the quote and be serviced or ask for another quote with a different quantity or service level. Both systems involve significant preparation and a certain amount of gamesmanship by the customer which greatly restrict their feasibility. [Nielsen, 1970 and EDP Analyzer, 1973]

A more conservative approach is to negotiate a contractual rate for large quantities of service at a given level of service over an extended period of time. Computer center management is provided a clearer forecast of requirements over a longer term. The customer is provided with a degree of certainty of service and awarded a rebate or volume

discount for assuming some of the computer center's risk. [Nielsen, 1968]

The flexible pricing technique is a realistic attempt at decentralization so that various levels in the organization can make the decisions for which they are best qualified. It reduces the suboptimization that results when projects are scheduled for the convenience of the computer center rather than the customer and vice versa.

A top level group can determine the general level of resource availability and utilization. Flexible pricing can provide data for the policy making group to improve its decision making. When prices fall too much below cost, this indicates over capacity. If prices rise too far above cost, then additional capacity should be considered. The final expected utilization should be the basis for establishing the price schedule which includes newly added capacity. Initial operating losses as a result of excess capacity should be budgeted as part of the cost of installing the new resource.

When an equitable budget constraint and an appropriate flexible price schedule have been established, the organization will benefit by having provided a mechanism and an incentive for the most knowledgeable persons, the customers, to make the detailed computer resource allocation decisions. [Smidt, Jun and Fall 1968, Nielsen, 1968 and 1970]

5. Currency Utilized--Hard or Soft Money

A reasonably reliable method is required to evaluate the intensity of a customer's desire to obtain service. Unpleasant activities, such as waiting, political capital expended, enemies made, effort expended to curry favor with computer center personnel and payment of money, may be considered a price. Money is currently the favored medium of exchange in the United States economy since resources may be transferred without being expended or consumed. [Sharpe, 1969]

In general terms, hard money may be defined as that which can be spent for any purpose and is effectively usable anywhere. Soft money may be defined as that which can be used only in some limited way or for a limited range of goods. Since the adjectives hard and soft also apply to budgetary allocations, additional clarification is desirable. Hard money in the form of cash, drawing rights or obligational authority is the type of money utilized when all customers receive their budgets totals entirely from the central organization without restrictions on how the customers allocate their budgets. When all customers receive their budget allocations from the central organization and an amount is identified which can be allocated only for data processing service, the earmarked portion is considered soft money and the balance of the budget which is not specifically identified is considered hard money. Therefore, it is possible to have a situation where cash dollars are viewed as soft money by a customer and hard money by the organization. [Smidt, Fall 1968]

Either hard or soft money can be utilized to pay for data processing services, provided that customers are subject to an effective budget constraint. Customers must clearly recognize that they are paying a price for the use of the data processing resource and that the supply of money is limited.

The size of a customer's budget reflects a judgement by the central organization about the relative importance of the goals the customer is attempting to achieve. This is particularly true for governmental organizations which have goals whose achievement cannot be easily measured in monetary terms.

An organization's routine budgeting procedures should provide that the increase in satisfaction or utility derived from each additional dollar allocated to a particular customer's budget is the same for all customers. This is especially important when the organization utilizes hard and soft money because of Gresham's Law. It states that if two currencies are used, and both are decreed to be of equal value but are actually of differing intrinsic values, the dearer currency will be hoarded and the cheaper one spent. Therefore, if the budgeting system does not provide equality of utility between customers and currencies, customers will tend to use computer resources as substitutes for other resources they otherwise would have to pay hard money for. [Smidt, Fall 1968, and EDP Analyzer, 1973]

Under a soft money system, the currency denomination reflected in the price schedule could be called CU's (computer units) instead of dollars. The data processing resource can be positively motivated to compete for business when its funding from the central organization is dependent upon the amount of CU's earned from its customers.

If CU's are the type of soft money utilized, they must be scarce and "inflation" must not be permitted. If a customer runs out of CU's, he must not be given more unless an equal number of CU's are reallocated from another customer or an organizational reserve. The total amount of CU's in the system cannot exceed what the data processing resource can satisfy. [Nielsen, 1970]

Hard money is the medium of exchange utilized when the organization sells data processing services to external customers. The internal price schedule serves to recover costs and allocate computer usage. The external price schedule also serves an income function for the selling organization. As a result, the internal and external price schedules are not necessarily the same.

Earlier discussion of microeconomic theory indicated that it is appropriate for computer services which are sold to internal customers to be priced at their marginal cost to the organization. Also, because of economies of scale, when prices are set equal to marginal cost, the total revenue to the organization will be less than its total cost. Therefore, prices to external customers should be above marginal cost.

Thus external customers tend to reduce the deficit from operating the data processing resource and increase the amount the organization has available for internally funded customers. In general, the less elastic an external customer's demand curve, the greater should be the discrepancy between the internal and external price schedules. [Smidt, Fall 1968, and Sharpe, 1969]

VI. PRICING OF U. S. NAVY DATA PROCESSING RESOURCES AN EXAMPLE

As of 30 June 1973, the U. S. Navy had 1,140 general purpose digital computers installed. Excluded are analog and digital computers which are built or modified to special government design and integral to a weapons or space system. [GSA, 1973] Individual organizations which physically operate this hardware are reporting cost and utilization data in accordance with a variety of instructions.

NAVSUP (Naval Supply Systems Command) requires monthly submission of an ADPE Costs and Utilization Report by all activities under its command. The objectives of this reporting system are to control allocation of funding, evaluate operating efficiency, control software development and maintenance, assist in effective management of resources and to serve as a basis for reports to the Chief of Naval Operations and the Secretary of the Navy.

The NAVSUP ADPE report does not directly compare cost with utilization or a measure of work. Hours of utilization and rental and operating costs are summarized, but no attempt is made to calculate average cost or to price CRU's or MU's. Purchased equipment is reported as a one-time charge and is not prorated over the hardware life.

A major NAVSUP field activity, which has several computer systems installed, prepares a quarterly price schedule of Data Processing Standard Rates for its computing and software

services. Charges for computing services are broken down into cost pools and are based on system or machine hours. Since monthly cost data is provided to in-house customers for information only and the price schedule does not discriminate between service levels, customers are encouraged to obtain all the service they can, as soon as they can.

Resource allocation is accomplished by formal and informal administrative regulations which include a priority system. Customers have no incentive to make dollar cost versus value of service trade-offs. However, they are frequently required to evaluate the adverse impact on project turnaround that may result if additional projects are requested or to decide whether to run project A or project B this week because only so much machine time is available. Therefore, delays in turnaround time act as an implicit rationing system which forces customers to make some trade-off decisions among their own projects. Decisions regarding the importance of customer A's project versus customer B's project to the organization, hopefully, is reflected in the priority system.

VII. CONCLUSIONS

Allocation of the data processing resource is a very complex problem. The details of the allocation problem and possible solutions will vary from organization to organization and will change with time. The basic concepts of economics and accounting, however, are of relevance and value to all.

The no charge-back approach may be a good arrangement for the user if the data processing resource is capable of satisfying all the user's desires. This is unlikely to occur in the long run if the organization has effective financial controls. Eventually, someone will be required to justify the indiscriminate growth of expenditures for data processing services.

The charge-back approach, in its various forms, is superior to the no charge-back approach, not only because users are placed in the role of customers and thus made aware of the fact that the data processing resource is not a free good, but also because it can lead to more effective utilization of scarce resources.

Allocation can be significantly improved by using two levels of decision making. A top-level policy group determines the general level of resource usage by various projects. Based on the established price schedule, the customer can make the detailed decisions about the required mix of computer services. He does this in light of his desires which are tempered by a budget constraint. Although each customer will

pursue his own self-interest, his actions will move the data processing center toward providing the appropriate mix of services which will mesh with the overall goals and priorities of the organization.

It must be noted, however, that establishing just any price schedule and some charge-back procedures will not accomplish the desired results. The pricing goals, principles and structure must be tailored to the individual organization after careful evaluation of its hardware and software capabilities, funding mix, organizational relationships and management's desires.

The average cost charge-back technique provides for full cost recovery and can minimize the indiscriminate utilization of the data processing resource if the customer is required to actually pay hard or soft dollars for services received, assuming of course that dollars are viewed as scarce resources by the customer.

The main undesirable aspect of the average cost technique is that it does a poor job of resource allocation. This is because, when utilization is high, prices are low, which encourages increased demand. The reverse incentive results when utilization is low. These results can be mitigated somewhat by introduction of additional variables in the price schedule to reflect quality of service. These include shift differentials, and pricing devices separately if they have different performance characteristics (e.g., storage devices-price disk packs and drum storage separately).

The flexible pricing charge-back technique is a more sophisticated approach towards cost recovery and resource allocation. Flexible pricing significantly enhances the resource allocation results because prices are high when demand is large relative to capacity and low when demand is small relative to capacity. The customer is provided not just an array of prices but, an array of flexible or adjustable prices that are responsive to demand. Both the no charge-back and the charge-back approaches require administrative regulations to varying degrees. Under flexible pricing, however, these regulations are reflected in the price schedule which improves decision making for the good of the entire organization.

Management should look beyond charging users simply as a means of cost recovery. The no charge-back approach, the average cost and flat rate techniques do not facilitate the most efficient use of resources; the flexible pricing technique does because it fully recognizes the physical and temporal aspects of data processing resource usage.

VIII. RECOMMENDATIONS

Although the literature concentrates on charge-back programs for commercial and university computing systems, military organizations should consider their use as well. Military organizations have the mission and deadline requirements of a commercial enterprise and experience the "free good" aspect of computer services in the university environment.

One article in the literature flatly states, "We have come across no good reason why flexible pricing cannot work in the average data processing installation." [EDP Analyzer, 1973] While charge-back systems, flexible pricing in particular, can provide significant benefits to the organization, customers and computer center management, no one technique is without its shortcomings. Therefore, it is recommended that every organization with a data processing facility carefully evaluate the implementation of a charge-back system.

Implementation of a charge-back system in a university environment where computer usage is presently a "free good" can alter resource utilization and strengthen the learning process. Without a charge-back system, students may be at liberty to continually rerun their projects in anticipation of running successfully this time. A charge-back system encourages the student to study a programming language more earnestly before attempting a project rerun and trains him

to work under a budget constraint which is an important aspect of his real life career.

Effective implementation of a charging system, at a university or in any other organization will involve a great deal of planning, communication and training. Therefore, it is recommended that actual implementation be phased in gradually.

The first phase after design and development of an average cost program might be to use the program to inform customers what their charges would be, without actually charging them. The second phase would be to actually start charging customers. Next, include shift differentials and additional resource prices in the price schedule. Implementation of a flexible pricing system could then be started after the organization has reasonably adjusted to the initial charge-back system. The time period for the transition from no charge-back to flexible pricing could take from six months to three or four years, depending on the organization.

The expected initial reaction to any charge-back system in a military organization is that "these projects must be processed at this level of service regardless of cost." That may very well be almost true and where it is considered so, establish contractual agreements for those projects which command such handling. Use of flexible pricing for the many other projects, however, can significantly improve the value of the work performed.

The subject of pricing and charging for data processing usage is one which has a large potential payoff in terms of improved efficiency and effectiveness of operations. Study of theory, practice and the literature identified in the Bibliography can greatly assist in designing a charge-back system that meets an organization's particular needs.

APPENDIX A

MICROECONOMICS CHARTS

FIGURE 1

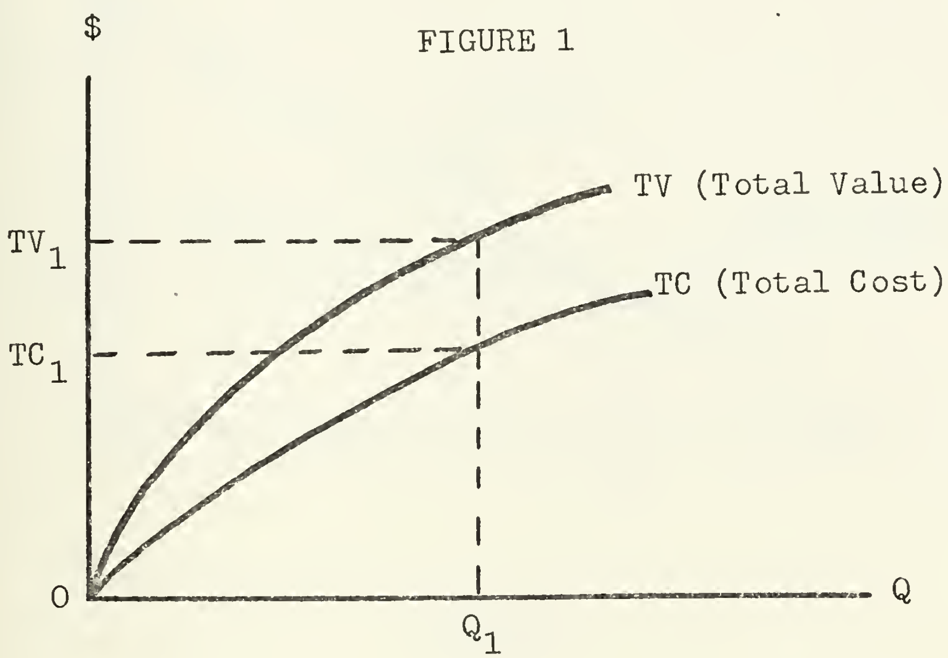


FIGURE 2

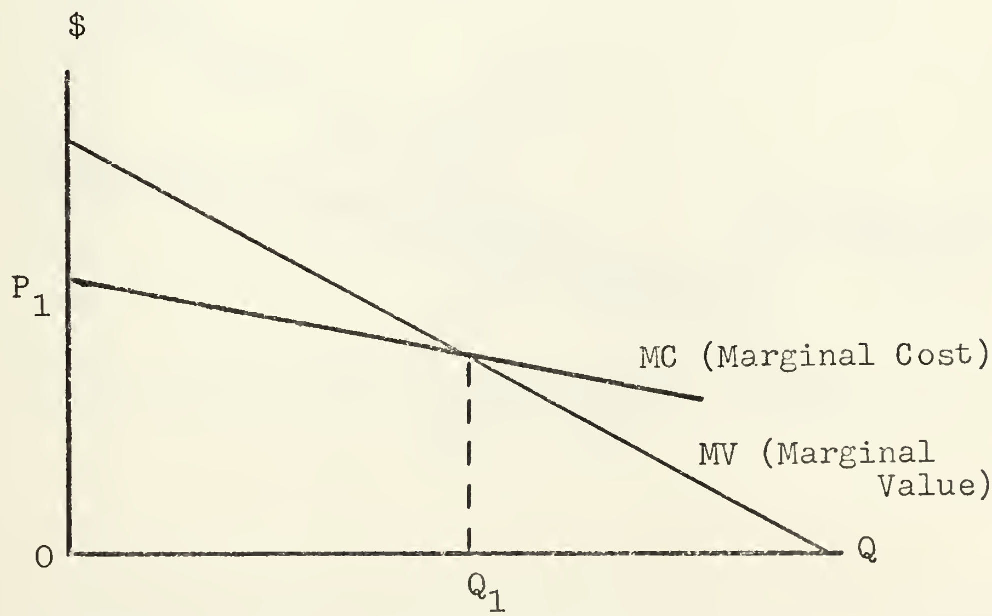


FIGURE 3

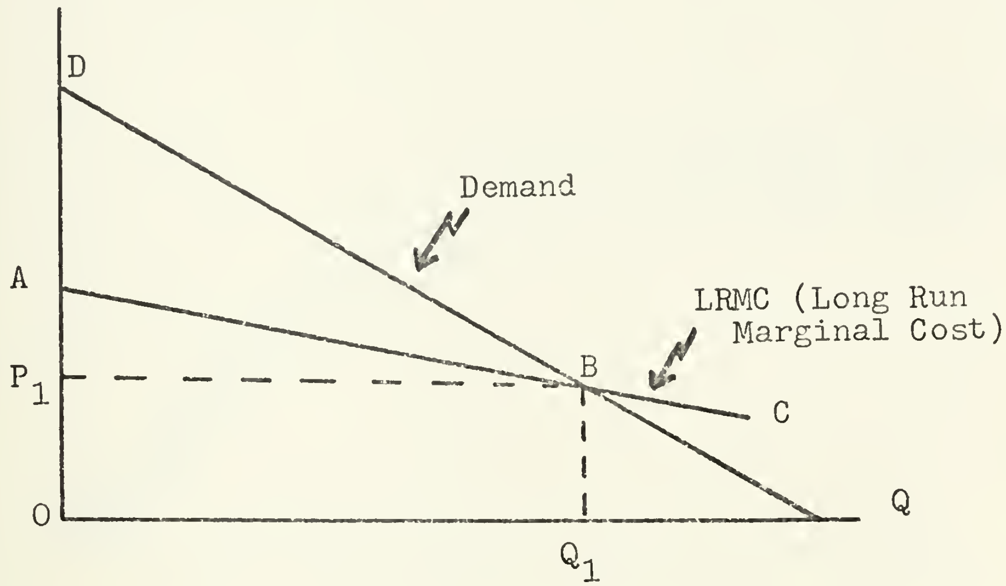


FIGURE 4

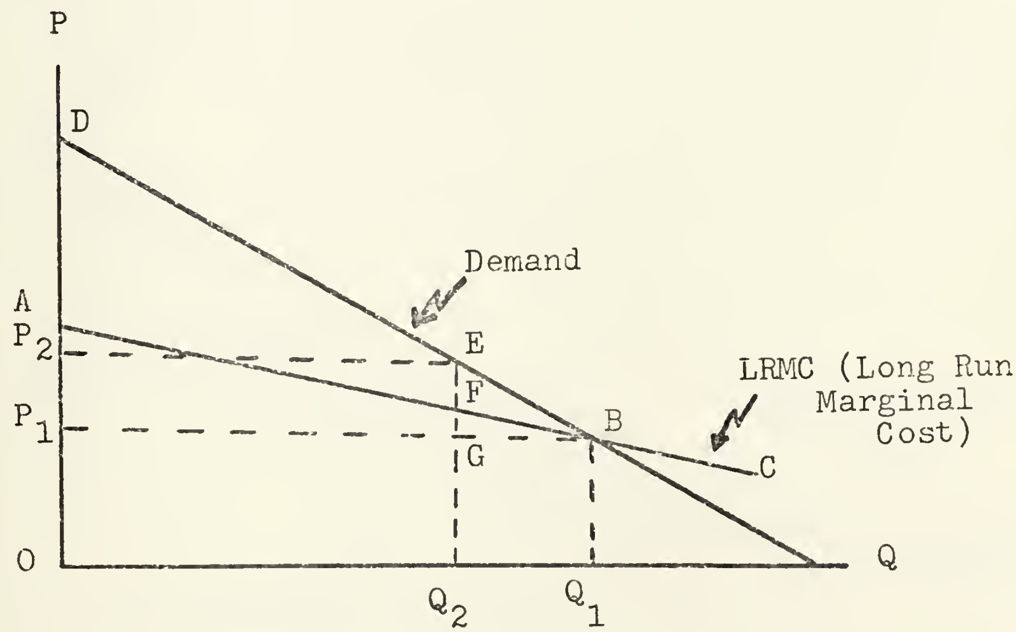


FIGURE 5

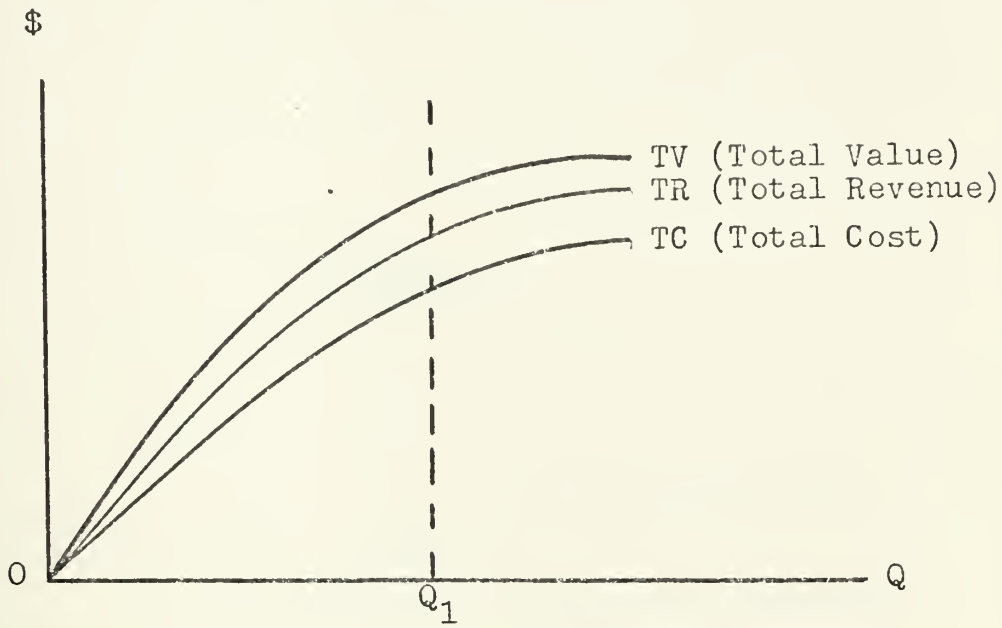


FIGURE 6

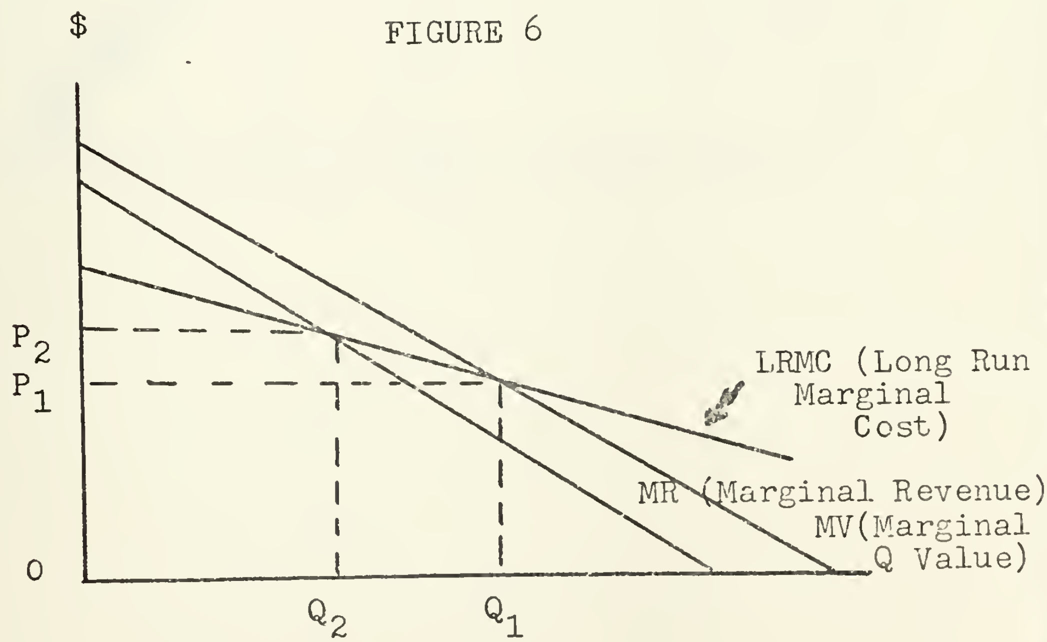
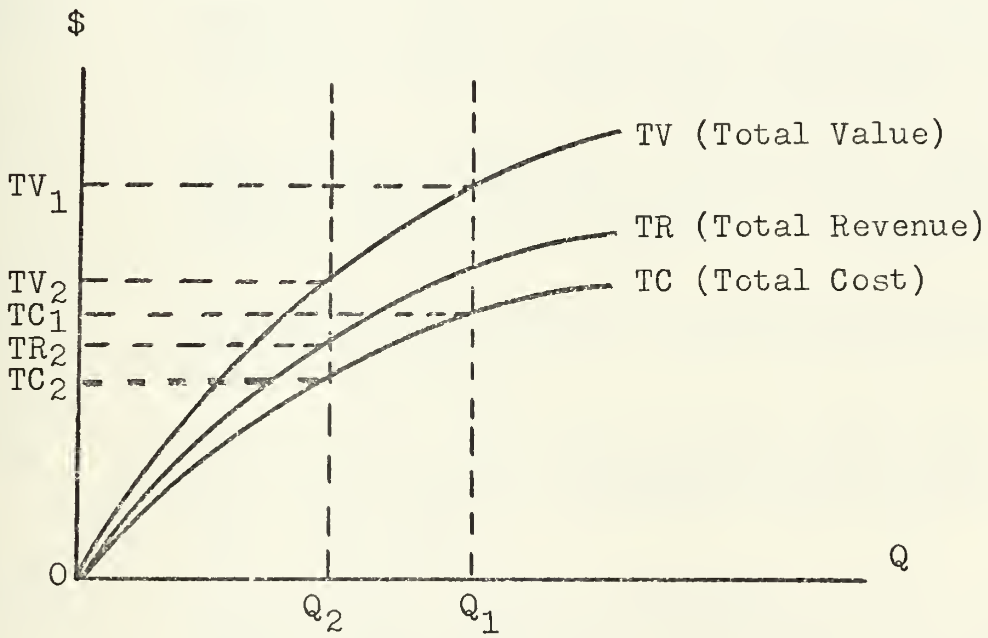


FIGURE 7


$
TV (Total Value)
TR (Total Revenue)
TC (Total Cost)
TV_1
TV_2
TC_1
TR_2
TC_2
0
Q
Q_2
Q_1

APPENDIX B

RESOURCE UTILIZATION MEASURES

TABLE I Possible Utilization Measures

ET:	Elapsed Time, the time from the start to the finish of a job (seconds).
CT:	Compute Time, the actual CPU time used by a job (seconds).
MT:	Memory Time, the integral over time of memory space occupied by a job (page seconds).
PMT:	Productive Memory Time, the integral over time of memory space occupied by a job while CPU cycles were being used (page seconds).
SS:	Swap Space, the integral over time of secondary storage space used by a job for swap image storage (or overlay storage, or virtual memory storage) (page seconds).
ST:	Swap Time, the amount of time spent swapping (overlaying or paging) a job in and out of main memory (seconds).
IOT_d:	Input Output Time, the sum of a job's I/O times for file references employing device $_d$ (tape drive, disk unit, device controller, I/O channel, PPU, etc.) (seconds).
ION_d:	Input Output Number, the number of I/O operations initiated by a job for file references employing device $_d$ (operations).
CR:	Cards Read, the number of input cards for a job (cards).
CP:	Cards Punched, the number of output cards for a job (cards).
LP:	Lines Printed, the number of output lines for a job (lines).
PT:	Plotting Time, the time from the start to the finish of a job's plot (seconds).
SU_c:	Set Ups (of tape reels, disk packs, special forms, etc.) required by a job on devices of class $_c$ (set ups).

FS_f: File Storage, the integral over time of storage space for a user's files on unit $_f$ (page days).

TR: Tape Reels, the integral over time of the number of reels reserved for a user (reel days).

DP: Disk Packs, the integral over time of the number of packs reserved for a user (pack days).

"A list of potential measures is shown in Table I; an example of a possible measurement unit for each item is shown in brackets. It should be noted that these measures are for a generalized time-sharing or multiprogramming system with a hierarchical secondary or file storage organization.

The first four measures are concerned with a job's utilization of the "main system." Elapsed time measures the period during which a job is in some sense "on" the system, while compute time indicates the period for which the CPU was employed. The memory measures are concerned with both the extent and the duration of memory occupancy. Although the page is used to illustrate the units of space, any other unit could be employed equally well.

The next eight measures are concerned with the I/O operations associated with the processing of a particular job. Swap space measures the amount of space required over time for the storage of program images. This would be applicable only in a system in which a job need not be totally core resident for the duration of its elapsed time (i.e., a system employing roll-in roll-out, swapping, paging, dynamic overlaying, or other such technique). This is also true

for the swap time measure, which is merely an indication of the amount of information swapped or otherwise transferred in and out of the main memory.

The other six I/O measures are concerned with the amount of I/O caused directly by a job (i.e., file references). Input output time and number provide measures of a job's usage of the various peripheral devices, control units, channels, peripheral processing units, etc. The measures in terms of cards and lines provide an indication of productive work, independent of the speed of the particular device and independent of operator interventions for card jams, paper breaks, etc.

The set ups measures provide an indication of the special work which was required on the part of the operations staff in order to enable a job to be processed. This would encompass the mounting of reserved tapes, the mounting of private disk packs, and the loading of special forms on a printer, etc.

The last three measures in the table are concerned not so much with the servicing of a job as they are with the servicing of a user over a period of time. Thus, file storage, tape reels and disk packs measure services which are provided to a user independent of the particular jobs he may have processed. In principle all of these measures are integrals over time. However, it is more likely to be practical to inventory a user's holdings on a daily or other periodic basis." [Nielsen, 1968]

TABLE II Possible Relations Between Measures and Resources

Elapsed Time:	terminals, transmission control units, telephone switching gear, terminal buffers in memory, any other portions of the system which are tied up for the duration of the life of a job
Compute Time:	central processing unit
Memory Time or Productive Memory Time:	high speed memory
Swap Space:	drums or highest speed disks
Swap Time:	drum (or disk) control units, swap channels, CPU cycles stolen
Input Output Time or Input Output Number:	tape units, disk drives, device control (units), I/O channels, PPU's etc.
Cards Read:	card reader, reader control unit, I/O channel
Cards Punched:	card punch, punch control unit, I/O channel
Lines Printed:	printer, printer control unit, I/O channel
Plotting Time:	plotter, plotter control unit, I/O channel
Set Ups:	operations staff, any system facilities disabled during the set up period
File Storage:	disks or other units on which on-line files are maintained
Tape Reels:	tapes storage facilities for tapes
Disk Packs:	packs, storage facilities for packs

Table II illustrates how resources in a time-shared or multiprogrammed system might be related to the various measures listed in Table I.

"It should be noted though that for any particular system these assignments would be modified in response to the unique characteristics of that system. Consider, for example, some of the alternatives with respect to memory usage. If it is possible for a job doing a great deal of I/O and very little computing to tie up memory space during all of the I/O transmissions, it would be preferable to use memory time as the usage measure. On the other hand, if it is possible for a job to be stranded in memory for long periods of time because other jobs do not relinquish the CPU, it would be preferable to use productive memory time as a measure. If the multiprogramming algorithm is such that one job is always given first priority for the CPU, it might be desirable to use memory time as a measure for that job and productive memory time as a measure for the other jobs in the system.

The use of swap space and swap time also depends upon the particular system used. If every job is given a fixed amount of drum space, independent of the actual amount of space required, drum space could just as well be related to the elapsed time measure. If the user does not have substantial control over the volume of data which is swapped the drum, control unit and channel resources could be related to swap space or elapsed time." [Nielsen, 1968]

BIBLIOGRAPHY

Anthony, R. N., Dearden, J., and Vancil, R. F., Management Control Systems, revised ed., p. 279-280, R. D. Irwin, 1972.

Bierman, H., Bonini, C. P., and Hausman, W. H., Quantitative Analysis for Business Decisions, 3rd ed., p. 366, R. D. Irwon, 1969.

Bray, T. J., "Rate Test Affects Vermonters' Habits, Formula Could Reduce Woes of Utilities," Wall Street Journal, v. 92, p. 28, 1 April 1975.

"Charging for Computer Services," EDP Analyzer, v. 12, p. 5-13, July 1974.

Dearden, J. and Nolan, R. L., "How to Control the Computer Resource," Harvard Business Review, v. 51, p. 68-78, November-December 1973.

Hootman, J. T., "The Pricing Dilemma," Datamation, v. 15, p. 61-66, August 1969.

Horngren, C., Cost Accounting: A Managerial Emphasis, 3rd ed., p. 405-407, Prentice-Hall, 1972.

Inventory of Automatic Data Processing Equipment in the United States Government for Fiscal Year 1973, p. 2-10, General Services Administration, 30 June, 1973.

Kanter, H., Moore, A., and Singer, N., "The Allocation of Computer Time by University Computer Centers," The Journal of Business, v. 41, p. 375-384, July 1968.

McLaughlin, R. A., "A Survey of 1974 DP Budgets," Datamation, v. 20, p. 52-56, February 1974.

McLaughlin, R. A., "1975 DP Budgets," Datamation, v. 21, p. 63-74, March 1975.

McLean, E. R., "Assessing Returns from the Data Processing Investment," Effective Vs Efficient Computing, p. 12-25, Informatics, Prentice-Hall, 1973.

Mobley, J. S., Pacific Telephone Company Computer Cost Allocation, paper presented to Professor Parry, Naval Postgraduate School, December 1974.

Nielsen, N. R., "Flexible Pricing: An Approach to the Allocation of Computer Resources," Proceedings AFIPS 1968 Fall Joint Computer Conference, v. 33, p. 521-531, MDI Publications, 1968.

Nielsen, N. R., "The Allocation of Computer Resources--Is Pricing the Answer?" Communications of the ACM, v. 13, p. 467-474, August 1970.

"Onrush of the Robots," Fortune, v. 69, p. 98, March 1964.

Sharpe, W. F., The Economics of Computers, p. 27-28, 455-465, Columbia University Press, 1969.

Singer, N., Kanter, H., and Moore, A., "Prices and the Allocation of Computer Time," Proceedings AFIPS 1968 Fall Joint Computer Conference, v. 33, p. 493-498, MDI Publications, 1968.

Smidt, S., "Flexible Pricing of Computer Services," Management Science, v. 14, p. B-581-600, June 1968.

Smidt, S., "The Use of Hard and Soft Money Budgets, and Prices to Limit Demand for Centralized Computer Facility," Proceedings AFIPS 1968 Fall Joint Computer Conference, v. 33, p. 499-503, MDI Publications, 1968.

Sobczak, J. J., "Pricing Computer Usage," Datamation, v. 20, p. 61-64, February 1974.

"The Effects of Charge-Back Policies," EDP Analyzer, v. 11, p. 1-14, November 1973.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55So Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	2
4.	ENS. G. G. Brown, Code 55Zr Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	2
5.	CAPT. O. B. Ross, Code 002 Comptroller Naval Postgraduate School Monterey, California 93940	1
6.	CAPT. S. R. Ruth, Code 041 Naval Supply Systems Command Headquarters Washington, D. C. 20376	1
7.	Dr. D. G. Williams, Code 0211 Director W. R. Church Computer Center Naval Postgraduate School Monterey, California 93940	1
8.	LCDR. T. R. Boyd P. O. Box 99118 San Diego, California 92109	1
9.	LCDR. W. F. Habermann Naval Air Station Box 21 FPO San Francisco 96654	1